COSMO, un modèle bayésien de la communication parlée : application à la perception des syllabes

Raphaël Laurent, Jean-Luc Schwartz, Pierre Bessière, Julien Diard

COSMO, un modèle bayésien de la communication parlée : application à la perception des syllabes

Raphaël Laurent[1, 2] Jean-Luc Schwartz[2] Pierre Bessière[1, 3] Julien Diard[4]
(1) LIG, UMR 5217 CNRS - Université de Grenoble
(2) GIPSA-Lab, Département Parole et Cognition (ICP), UMR 5216 CNRS - Université de Grenoble
(3) LPPA, UMR 7152 CNRS - Collège de France, Paris
(4) LPNC, UMR 5105 CNRS - Université de Grenoble
Raphael.Laurent@Gipsa-lab.Grenoble-inp.fr

RÉSUMÉ
Le travail présenté ici s'inscrit dans le cadre de la modélisation computationnelle pour comparer les théories motrice, auditive, et sensorimotrice de la communication parlée. Plus précisément, nous définissons un modèle bayésien d'agent communicant et le simulons pour réaliser des tâches de perception de syllabes suivant ces différentes théories. Les résultats de nos simulations s'intègrent dans un cadre général selon lequel les théories motrices de la perception sont plus robustes au bruit, et les théories auditives favorisées par les non-linéarités.

ABSTRACT
COSMO, a Bayesian model of speech communication, applied to syllable perception

This works uses the computational modeling framework to compare motor, auditory, and sensorimotor theories of speech communication. More precisely, we define a Bayesian model of a communicating agent which we simulate to carry out syllable perception tasks according to these theories. Our simulation results are consistent with the idea that motor theories of speech perception are more robust to noise, and that auditory ones are favored by non-linearities.

MOTS-CLÉS : Perception de la parole, programmation bayésienne, modélisation cognitive.

KEYWORDS: Speech perception, Bayesian programming, cognitive modeling.

Introduction

Une question centrale dans le domaine de la parole concerne la nature des représentations et des processus cognitifs qui interviennent dans la communication. Trois grands groupes de théories sont au coeur de ce débat classique : les théories motrices, auditives, et sensorimotrices. Parmi les principaux arguments, qui reposent sur des données expérimentales sur la variabilité et l'invariance, on peut citer le phénomène de coarticulation en faveur des théories motrices (Galantucci *et al.*, 2006) ou le principe d'équivalence motrice, en faveur des théories auditives (Guenther *et al.*, 1998; Diehl *et al.*, 2004). Par ailleurs, des théories sensorimotrices marient ces approches (Guenther, 1995) ; par exemple la théorie de la perception pour le contrôle de l'action, *PACT* (Schwartz *et al.*, 2010), défend la co-structuration du système auditif et du système moteur.

Pourtant, l'observation isolée de ces propriétés ne rend pas les arguments associés suffisamment

décisifs pour trancher, et le débat théorique stagne. Nous pensons que la modélisation computationelle peut apporter un éclairage supplémentaire car elle permet la comparaison efficace et systématique de ces théories et de leurs propriétés.

La Programmation bayésienne procure un cadre mathématique permettant de telles comparaisons, dans lequel le même outil, à savoir les probabilités, est utilisé à la fois pour définir les modèles et pour les comparer. Forts de cet outil, nous adoptons une approche intégrative, permettant de regrouper les théories motrice, auditive, et sensorimotrice au sein d'un unique modèle bayésien unificateur. Cela rend possible des études systématiques de ces théories, ce que nous faisons avec des tests quantitatifs.

Suivant cette approche, nous avons, dans des travaux précédents, conçu et implémenté un modèle bayésien d'agent communiquant, basé sur l'internalisation de la situation de communication. Nous présentons ici une extension de ce modèle permettant d'étudier les syllabes.

Dans ce qui suit nous commençons par rappeler le modèle bayésien d'agent communicant que nous proposons, et les premiers résultats qu'il a permis d'obtenir. Ensuite, nous montrons comment ce modèle général peut être étendu, pour traiter le cas des syllabes. Enfin, nos simulations montrent que les théories motrices disposent d'une meilleure capacité de généralisation des apprentissages, et confirment également la supériorité du modèle moteur dans les cas linéaires.

1 *COSMO* : un modèle bayésien d'agent communiquant

1.1 Définition du modèle *COSMO*

Dans des travaux précédents (Moulin-Frier *et al.*, 2012), nous avons conçu un modèle bayésien d'agent communiquant, que nous baptisons *COSMO*, pour *Communicating about Objects using SensoriMotor Operations*.

Ce modèle provient de la modélisation de la situation de communication (Fig. 1) : deux agents veulent communiquer à propos d'un objet de l'environnement. L'agent Speaker (locuteur), pour désigner l'objet O^S, réalise un geste moteur M qui produit un son S permettant au Listener (auditeur) de reconnaître l'objet O^L. Un mécanisme d'attention partagée (par exemple la deixis) permet de valider le succès de la communication (variable C_{Env}). Le modèle *COSMO*, dont l'acronyme reprend les variables qui viennent d'être présentées, est basé sur l'hypothèse fondamentale que cette situation de communication peut être internalisée et émulée dans le cerveau de chaque agent (Fig. 1), qui est alors en mesure d'agir aussi bien en tant que locuteur qu'auditeur.

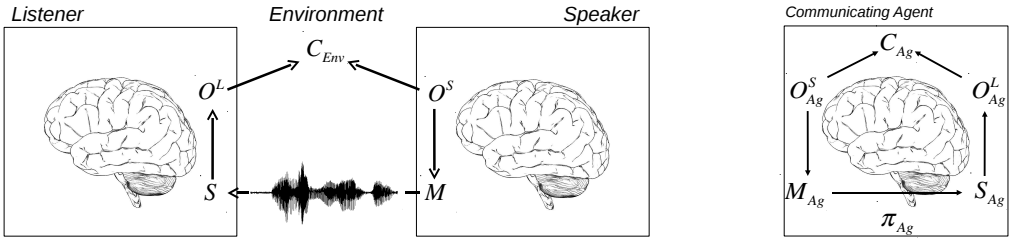

FIGURE 1: À gauche, le modèle de la situation de communication ; à droite, le modèle *COSMO* d'agent communicant basé sur l'internalisation de la situation de communication.

Bien que notre modèle soit compatible avec des définitions au sens très large du terme “objet”, dans le cadre du travail présenté ici, ce terme désignera simplement des unités phonologiques partagées par le locuteur et l'auditeur.

Notre modèle d'agent, noté π_{Ag}, est entièrement décrit par la distribution de probabilité conjointe sur l'ensemble de nos variables. Nous choisissons de la décomposer de la manière suivante.

$$\begin{aligned} & P(O^S_{Ag}\ M_{Ag}\ S_{Ag}\ O^L_{Ag}\ C_{Ag} \mid \pi_{Ag}) \\ = \quad & P(O^S_{Ag} \mid \pi_{Ag}) \times P(M_{Ag} \mid O^S_{Ag}\ \pi_{Ag}) \times P(S_{Ag} \mid M_{Ag}\ \pi_{Ag}) \times \\ & P(O^L_{Ag} \mid S_{Ag}\ \pi_{Ag}) \times P(C_{Ag} \mid O^S_{Ag}\ O^L_{Ag}\ \pi_{Ag})\ . \end{aligned}$$

Le système moteur $P(M_{Ag} \mid O^S_{Ag}\ \pi_{Ag})$ décrit la connaissance qu'a l'agent sur les gestes moteurs associés aux objets.

Le système sensorimoteur $P(S_{Ag} \mid M_{Ag}\ \pi_{Ag})$ décrit la connaissance qu'à l'agent sur la relation entre le geste articulatoire et sa conséquence sensorielle.

Le système auditif $P(O^L_{Ag} \mid S_{Ag}\ \pi_{Ag})$ décrit la connaissance qu'a l'agent sur la relation entre les stimuli et les objets.

Le système de validation de la communication $P(C_{Ag} \mid O^S_{Ag}\ O^L_{Ag}\ \pi_{Ag})$ décrit la connaissance qu'a l'agent sur le succès de la communication. La communication est un succès lorsque les objets considérés du point de vue du locuteur et de l'auditeur sont les mêmes ($O^S_{Ag} = O^L_{Ag}$).

Le prior sur les objets $P(O^S_{Ag} \mid \pi_{Ag})$ décrit la connaissance a priori qu'a l'agent sur la répartition des objets dans l'environnement.

1.2 Résultats théoriques et expérimentaux

Ce travail de modélisation nous a conduit à trois résultats principaux (Moulin-Frier *et al.*, 2012).

Premièrement, notre modèle d'agent sensorimoteur, qui est capable de mener à bien les tâches de production ainsi que de perception dans le cadre de chacune de nos trois grandes théories motrice, auditive, et sensorimotrice, permet de les comparer de manière systématique.

Deuxièmement, nous avons démontré théoriquement que, sous deux hypothèses principales, les théories auditive et motrice de la perception sont indistingables. La première hypothèse est que le système auditif est appris sous la supervision d'un agent maître ayant les mêmes prototypes moteurs que l'agent apprenant. La seconde hypothèse est que l'agent apprenant a parfaitement identifié les propriétés de l'environnement. Sous ces hypothèses, nous avons montré que les réponses auditive et motrice aux tâches de perception sont rigoureusement identiques.

Troisièmement, en simulant des tâches de reconnaissance de voyelles, nous avons montré que les théories motrices sont plus robustes aux perturbations testées (bruit d'environnement, différences entre locuteurs) mais que les théories auditives sont favorisées par la présence de non-linéarités dans la relation articulatori-acoustique (Stevens, 1972).

2 Extension du modèle *COSMO* au cas des syllabes

2.1 Définition du modèle *COSMO-Syllabes*

Le modèle *COSMO* qui vient d'être présenté a été défini et utilisé pour manipuler des objets de

type voyelle. Nous présentons maintenant *COSMO-Syllabes*, une extension de ce modèle dans laquelle les objets (O_S et O_L) correspondent aux syllabes /ba/, /bi/, /bu/, /ga/, /gi/, /gu/, /da/, /di/ et /du/, qui sont construites à partir des voyelles et des plosives les plus fréquentes.

Une syllabe se définit dans ce cadre simplifié comme une transition continue entre deux états : un état pour lequel le conduit vocal est presque clos, et un état dans lequel le conduit vocal s'est stabilisé dans une position plus ouverte. Dans ce qui suit, une syllabe est décrite par la donnée de ces deux états : un état consonne, et un état voyelle ; la trajectoire est négligée.

Dans le modèle *COSMO-Syllabes*, que nous notons π, la variable M du modèle *COSMO* se dédouble en M_V et M_C, les gestes moteurs respectivement de la voyelle et de la consonne. De même, la variable S se dédouble en S_V et S_C, les représentations sensorielles, respectivement de la voyelle et de la consonne. Les autres variables ne changent pas.

Certains termes de la distribution de probabilité conjointe (Fig. 2) sont également modifiés.

$$
\begin{aligned}
& P(O_S\ M_V\ M_C\ S_V\ S_C\ O_L\ C \mid \pi) \\
= \quad & P(O_S \mid \pi) \times \\
& P(M_V \mid O_S\ \pi) \times P(M_C \mid M_V\ O_S\ \pi) \times \\
& P(S_V \mid M_V\ \pi) \times P(S_C \mid M_C\ \pi) \times \\
& P(O_L \mid S_V\ S_C\ \pi) \times \\
& P(C \mid O_S\ O_L\ \pi)
\end{aligned}
$$

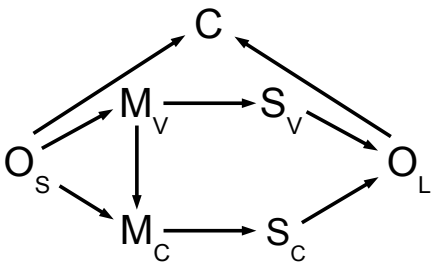

FIGURE 2: À gauche, la distribution de probabilité conjointe du modèle *COSMO-Syllabes* ; à droite, une représentation graphique de cette structure de dépendance probabiliste.

Le système moteur se décompose en deux termes : $P(M_V \mid O_S\ \pi)$ et $P(M_C \mid M_V\ O_S\ \pi)$, qui décrivent les connaissances sur les gestes moteurs permettant de produire la partie voyelle, respectivement consonne, de la syllabe.

Considérons le terme $P(M_C \mid M_V\ O_S\ \pi)$: le geste moteur de la partie consonne d'une syllabe est conditionné par le geste moteur de la partie voyelle de cette syllabe. Il s'agit d'un choix de modélisation : on raisonne suivant un scénario dans lequel la voyelle est anticipée au moment de produire la consonne, et influence ainsi cette production. Ce terme $P(M_C \mid M_V\ O_S\ \pi)$ traduit donc directement le phénomène de coarticulation en reprenant le classique modèle de "perturbation consonantique" (Öhman, 1966).

Le système sensorimoteur se décompose également en deux termes : $P(S_V \mid M_V\ \pi)$ et $P(S_C \mid M_C\ \pi)$, qui décrivent les connaissances de la relation entre gestes articulatoires et conséquences sensorielles pour la voyelle, et respectivement pour la consonne.

Le système auditif $P(O_L \mid S_V\ S_C\ \pi)$ décrit les connaissances sur la relation entre stimuli voyelle et consonne d'une part, et les objets d'autre part.

Le prior sur les objets et **le système de validation** de la communication sont inchangés.

2.2 Inférences probabilistes pour des tâches de perception

Dans le modèle *COSMO*, une tâche de perception s'exprime sous forme de question probabiliste posée au modèle Bayésien. Dans le modèle *COSMO-Syllabes*, une tâche de perception consiste à,

étant donné un stimulus (S_V, S_C), calculer la distribution de probabilité sur les objets. Nous nous limitons à en comparer deux versions, selon que l'on choisit comme pivot les objets considérés du point du locuteur (O_S), ou de l'auditeur (O_L).

Perception dans le cadre d'une théorie auditive :
Dans le cadre d'une théorie auditive, la question de perception s'écrit $P(O_L \mid S_V\ S_C\ \pi)$. Étant donné un stimulus syllabe, quelle est la distribution de probabilité sur les objets syllabes, envisagés d'un point de vue auditif ? Ce terme est présent directement sous cette forme dans la distribution de probabilité conjointe, il s'agit juste de le consulter.

Perception dans le cadre d'une théorie motrice :
Dans le cadre d'une théorie motrice, la question de perception s'écrit $P(O_S \mid S_V\ S_C\ \pi)$. Étant donné un stimulus syllabe, quelle est la distribution de probabilité sur les objets syllabes envisagés d'un point de vue moteur ? L'inférence bayésienne donne :

$$P(O_S \mid S_V\ S_C\ \pi) = \frac{1}{Z} \sum_{M_V} \left[P(M_V \mid O_S\ \pi) P(S_v \mid M_V\ \pi) \sum_{M_C} P(M_C \mid M_V\ O_S\ \pi) P(S_C \mid M_C\ \pi) \right] ,$$

où Z est une constante de normalisation.

Notre implémentation d'une théorie motrice de perception de la parole revient donc à faire de l'analyse par la synthèse : on parcourt l'ensemble des gestes articulatoires et on considère leur probabilité de correspondre aux stimuli (facteurs $P(S_V \mid M_V\ \pi)$ et $P(S_C \mid M_C\ \pi)$) et aux objets considérés (facteurs $P(M_V \mid O_S\ \pi)$ et $P(M_C \mid M_V\ O_S\ \pi)$). En particulier, avec cette approche bayésienne, le problème de la redondance dans l'inversion motrice ne se pose pas puisque tous les cas possibles sont pris en compte, pondérés par leur probabilité.

2.3 Génération de syllabes avec un modèle articulatoire réaliste

Pour pouvoir apprendre les paramètres de notre modèle dans un premier temps, puis l'évaluer ensuite, nous avons besoin de données articulatoires et acoustiques de syllabes. Pour générer ces données, nous utilisons un modèle réaliste de conduit vocal : *VLAM*, the *Variable Linear Articulatory Model* (Maeda, 1990). Ce modèle prend en compte sept paramètres articulatoires, décrivant la position de la mâchoire et du larynx, la forme de la langue et des lèvres, qui sont interprétables en termes de commandes phonétiques, et qui sont très proches de commandes musculaires (Maeda et Honda, 1994). L'aire de chacune des 28 sections du conduit vocal est estimée comme une combinaison linéaire de ces sept paramètres, ce qui permet ensuite de calculer la fonction de transfert et les formants (Badin et Fant, 1984).

Pour rendre les calculs d'inférence abordables, nous limitons le nombre de paramètres utilisés. Dans l'espace acoustique, on décrit les voyelles par les formants F_1 et F_2, et les consonnes par F_2 et F_3. Dans l'espace articulatoire, on décrit les voyelles par les trois paramètres `TB` (corps de la langue) `TD` (dos de la langue) et `LH` (écart entre les lèvres), et pour les consonnes on ajoute `J` (mâchoire) et `APEX` (pointe de la langue).

Nous décrivons maintenant le processus de génération de dictionnaires de syllabes, qui sont produits à partir de tirages gaussiens avec différentes variances autour de prototypes moteurs. Le dictionnaire de petite variance sera utilisé pour apprendre les paramètres du modèle *COSMO-Syllabes*, et les dictionnaires de plus grande variance seront utilisés pour tester ses capacités de généralisation.

En prenant comme cibles acoustiques les valeurs moyennes de formants des voyelles /a/, /i/, /u/ (Meunier, 2007), nous obtenons des valeurs prototypiques de nos paramètres articulatoires pour ces voyelles. En tirant selon une loi gaussienne de variance prédéfinie autour de chacune de ces valeurs prototypiques, nous produisons trois ensembles de 25000 points dans l'espace vocalique. Nous faisons l'hypothèse d'une coarticulation maximale : la consonne est formée à partir de la voyelle en ne mobilisant que deux articulateurs. D'une part `LH`, `TD` ou `APEX`, permet de produire respectivement un /b/, un /g/ ou un /d/ ; et d'autre part `J` garantit de la variabilité sur la consonne.

Nous montrons (Fig. 3) les syllabes du corpus de petite variance obtenu en suivant ce processus. Sans surprise, les similitudes classiques entre /bu/ et /gu/, et entre /di/ et /gi/ y sont présentes. Par ailleurs, les /da/ y sont assez proches des /ba/, ce qui s'explique par notre choix de produire des /d/ plus proches des dentales que des alvéolaires, comme c'est le cas en français (Schwartz *et al.*, 2012).

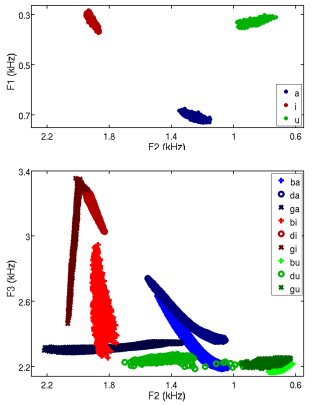

FIGURE 3: Les syllabes dans l'espace acoustique : projection de la voyelle dans (F_2, F_1) en haut, et de la consonne dans (F_2, F_3) en bas.

2.4 Apprentissage des paramètres du modèle bayésien

Les facteurs $P(M_V \mid O_S\ \pi)$ et $P(O_L \mid S_V\ S_C\ \pi)$ sont appris à partir des connaissances locales issues du corpus de syllabes de faible variance, le premier sous forme de lois de probabilité gaussiennes, et le second sous forme d'inversion de gaussiennes. En revanche, les facteurs $P(S_V \mid M_V\ \pi)$, $P(S_C \mid M_C\ \pi)$ et $P(M_C \mid M_V\ O_S\ \pi)$ font intervenir des connaissances globales, qui sont apprises sous forme de tables de probabilités discrètes. Plus précisément, nous faisons l'hypothèse que la transformation articulatori-acoustique est connue sur tout l'espace articulatoire (voyelle et consonne). Nous faisons de plus l'hypothèse que le lien entre la consonne (/b/ /d/ ou /g/) et l'articulateur à actionner (lèvres, pointe ou dos de la langue) est parfaitement identifié : le modèle sait produire ces consonnes à partir de toutes les configurations articulatoires de voyelle.

2.5 Simulations et résultats

Dans des travaux précédents, nous avons étudié l'effet de perturbations (bruit de l'environnement, différences entre locuteurs) sur les taux de reconnaissance prédits par les différentes théories. Ici, nous voulons tester la capacité de généralisation de nos modèles. Pour cela, nous utilisons des corpus de données syllabes avec des variances différentes : le corpus ayant la variance la plus petite sert à l'apprentissage, et les corpus de variances supérieures à l'évaluation.

Nous calculons les distributions de probabilité $P(O \mid S_V\ S_C\ \pi)$ à partir des stimuli des corpus de test, et comparons les syllabes reconnues avec les syllabes auxquelles ces stimuli correspondent. En calculant la moyenne de ces distributions sur l'ensemble des stimuli de chaque catégorie de syllabe, nous obtenons des matrices de confusion. Nous en montrons un exemple (Fig. 4) tiré de la simulation du modèle moteur sur les syllabes du corpus de faible variance.

	/ba/	/bi/	/bu/	/ga/	/gi/	/gu/	/da/	/di/	/du/
/ba/	0.817	0	0	0.005	0	0	0.178	0	0
/bi/	0	0.9995	0	0	0.0005	0	0	0	0
/bu/	0	0	0.654	0	0	0.346	0	0	0
/ga/	0.0053	0	0	0.9945	0	0	0.0002	0	0
/gi/	0	0.0027	0	0	0.9528	0	0	0.0445	0
/gu/	0	0	0.4469	0	0	0.5523	0	0	0.0008
/da/	0.1354	0	0	0.0002	0	0	0.8644	0	0
/di/	0	0.0009	0	0	0.1319	0	0	0.8672	0
/du/	0	0	0.0002	0	0	0.0055	0	0	0.9943

FIGURE 4: Confusions du modèle moteur de perception sur le corpus syllabes de faible variance. Chaque ligne, qui correspond à un type de stimuli, donne la probabilité des syllabes reconnues.

À partir de telles matrices de confusion, nous définissons trois scores globaux : les taux de reconnaissance de la voyelle de la syllabe (*RV*), de la consonne de la syllabe (*RC*), et de la syllabe en entier (*RS*). Nous montrons (Fig. 5) l'évolution de ces scores lorsque la variance du corpus d'évaluation (exprimée en pourcentage de la variance du corpus d'apprentissage) augmente.

variance	100%			175%			250%			325%			400%		
score	RV	RC	RS	RV	RC	RS	RV	RC	RS	RV	RC	RS	RV	RC	RS
modèle moteur	1	0.88	0.88	0.99	0.89	0.89	0.99	0.90	0.90	0.99	0.90	0.90	0.99	0.90	0.90
modèle auditif	0.99	0.90	0.90	0.98	0.89	0.89	0.95	0.87	0.85	0.91	0.84	0.82	0.88	0.81	0.78

FIGURE 5: Scores de reconnaissance des modèles issus des théories auditive et motrice.

Pour une faible variance des données du corpus d'évaluation, les deux modèles prédisent des scores de classification similaires, mais lorsque cette variance augmente, le modèle auditif voit ses performances baisser significativement, alors que le modèle moteur reste stable. Ces résultats confortent l'observation selon laquelle le modèle moteur est plus robuste en conditions dégradées.

2.6 Discussion

Nos résultats montrent un net avantage du modèle moteur qui, pensons-nous, doit être relativisé. En ce qui concerne les capacités de généralisation du modèle moteur, elles viennent sans doute du fait qu'il dispose de beaucoup de connaissances globales. Par exemple, notre modèle est capable de produire précisément les gestes articulatoires correspondant à n'importe quelle cible acoustique. Ce n'est pas le cas du locuteur naturel, dès que l'on s'éloigne des sons des langues qu'il maîtrise. Nous étudions dans des travaux en cours des scénarii d'apprentissage plus réalistes.

Par ailleurs, une partie des performances du modèle moteur s'explique par la linéarité. Les consonnes /b/ /d/ /g/ ne diffèrent que par le lieu d'articulation ; ajouter une distinction sur le mode d'articulation (entre plosives et fricatives) introduirait une non-linéarité qui ferait baisser les performances du modèle moteur, redonnant le dessus au modèle auditif.

Les résultats de simulations présentés dans cet article restent préliminaires, et la principale contribution est théorique : nous disposons maintenant d'un modèle unique qui rend abordable la simulation des tâches de production et de perception de syllabes, et qui permet d'étudier au même niveau les théories motrice, auditive, et sensorimotrice.

Pour le moment, nous nous intéressons surtout à la comparaison des théories auditive et motrice. Si, comme nous le pensons, les connaissances motrices et auditives apportent des informations

complémentaires selon les contextes, une théorie sensorimotrice doit se donner les moyens d'extraire l'information utile et d'en tirer parti. Dans le cadre de notre modèle bayésien, cela prend la forme d'une fusion de capteurs, un problème à part entière à étudier en tant que tel.

Conclusion

Nous avons présenté le modèle *COSMO* d'agent bayésien communicant, et les premiers résultats qu'il a permis d'obtenir sur les conditions d'indistingabilité des théories motrice et auditive, sur la robustesse du modèle moteur aux bruits, et sur la supériorité du modèle auditif en présence de non-linéarités. Nous avons montré comment ce modèle général peut être étendu en un modèle *COSMO-S* permettant d'étudier les syllabes. Enfin, nos simulations suggèrent que les théories motrices pourraient disposer d'une meilleure capacité de généralisation des apprentissages. Dans une optique de complémentarité entre système moteur et système auditif, nous essayons de montrer dans des travaux en cours comment, dans un modèle disposant d'un classifieur audio, l'apprentissage de connaissances motrices par imitation peut faire *émerger* la notion de consonne.

Références

Badin, P. et Fant, G. (1984). Notes on Vocal Tract Computation. *In Quarterly Progress and Status Report, Dept for Speech, Music and Hearing, KTH, Stockholm*, pages 53–108.

Diehl, R. L., Lotto, A. J. et Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55(1):149–179.

Galantucci, B., Fowler, C. A. et Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3):361–377.

Guenther, F. H. (1995). A modeling framework for speech motor development and kinematic articulator control. *In Proceedings XIIIth International Congress of Phonetic Sciences*, volume 2, page 92–99. Citeseer.

Guenther, F. H., Hampson, M. et Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105:611–633.

Maeda, S. (1990). Compensatory articulation during speech : Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. *In* Hardcastle, W. et Marchal, A., éditeurs : *Speech production and speech modeling*, pages 131–149. Kluwer Academic.

Maeda, S. et Honda, K. (1994). From EMG to formant patterns : the implication of vowel spaces. *Phonetica*, 51:17–29.

Meunier, C. (2007). Phonétique acoustique. *In* Auzou, P., éditeur : *Les dysarthries*, pages 164–173. Solal.

Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.-L. et Diard, J. (2012). Adverse conditions improve distinguishability of auditory, motor and percep-tuo-motor theories of speech perception : an exploratory Bayesian modeling study. *Language and Cognitive Processes*, page In Press.

Öhman, S. E. G. (1966). Coarticulation in vcv utterances : Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1):151–168.

Schwartz, J.-L., Basirat, A., Ménard, L. et Sato, M. (2010). The perception for action control theory (PACT) : a perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, pages 1–19.

Schwartz, J.-L., Boë, L.-J., Badin, P. et Sawallis, R., T. (2012). Grounding stop place systems in the perceptuo-motor substance of speech : On the universality of the labial-coronal-velar stop series. *Journal of Phonetics*, 40:20–36.

Stevens, K. (1972). The quantal nature of speech : Evidence from articulatory-acoustic data. *In* David, E. et Denes, P., éditeurs : *Human communication : A unified view*, pages 51–66. McGraw-Hill.